



**QUEEN'S
UNIVERSITY
BELFAST**

The Belfast Induced Natural Emotion Database

Sneddon, I., McRorie, M., McKeown, G., & Hanratty, J. (2012). The Belfast Induced Natural Emotion Database. *IEEE Transactions on Affective Computing*, 3(1), 32-41. [5975142]. <https://doi.org/10.1109/T-AFFC.2011.26>

Published in:

IEEE Transactions on Affective Computing

Document Version:

Early version, also known as pre-print

Queen's University Belfast - Research Portal:

[Link to publication record in Queen's University Belfast Research Portal](#)

General rights

Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

The Belfast Induced Natural Emotion Database

Ian Sneddon, Margaret McRorie, Gary McKeown and Jennifer Hanratty

Abstract—For many years psychological research on facial expression of emotion has relied heavily on a recognition paradigm based on posed static photographs. There is growing evidence that there may be fundamental differences between the expressions depicted in such stimuli and the emotional expressions present in everyday life. Affective computing, with its pragmatic emphasis on realism, needs examples of natural emotion. This paper describes a unique database containing recordings of mild to moderate emotionally coloured responses to a series of laboratory based emotion induction tasks. The recordings are accompanied by information on self-report of emotion and intensity, continuous trace-style ratings of valence and intensity, the sex of the participant, the sex of the experimenter, the active or passive nature of the induction task and it gives researchers the opportunity to compare expressions from people from more than one culture.

Index Terms—Natural Emotion, Database, Emotion Induction, Emotional Corpora, Affective Annotation.

1 INTRODUCTION

THEORIES about the capacity of humans to recognise emotions from the facial behaviour of others are based largely on the study of highly selected still photographs of posed expressions – an experimental paradigm that has been called the “standard method” [1]. Theories of emotional expression are, of course, attempts to explain natural, not posed human behaviour. Despite this fact, there has been relatively little research on natural dynamic examples of emotion (see [2] for a review). Although the lack of natural evidence in support of theories of emotional expression has been criticised [1], the reluctance to use examples of natural expression when studying emotion is at least partly understandable. When using posed photographs the researcher typically presents participants with examples of prototypical facial expressions based on the Facial Action Coding System (FACS) developed by Ekman and others [3]. Thus the recognition of the emotion can be judged against the ‘correct’ expression adopted by the poser. The use of examples of natural emotion is much more problematic. In this case the researcher is using a spontaneous facial expression with no means of knowing objectively what emotion is being experienced by the encoder or the extent to which the facial expression reflects that emotion. The problem can be conceived as a trade-off between ecological validity and reliability, with the decision over which approach to favour being influenced by the nature of the research question and its sensitivity to each of these factors. If natural everyday emotional expressions showed only quantitative differences from the posed examples (perhaps occurring as a weaker version of the prototypical forms) then perhaps the over-reliance on these examples would be

of little importance, but recent research suggests that there may be fundamental differences between the two. A number of research studies have indicated that even actors do not show the prototypical patterns of facial behaviour when attempting to convey emotion. It has been reported [4] that Hollywood actors only exhibit the prototypical patterns specified by Ekman and Friesen when portraying happiness, and more recent research [5] found none of the prototypical patterns of action units when experienced actors simulated strong emotions. Additionally, it has been suggested [6] that in speech, there may be important qualitative differences between posed and natural examples of emotion expression.

Recently the over-reliance on static images has also been questioned. Evidence is mounting which demonstrates that important information is contained in the dynamic unfolding of facial expressions over time [7], [8] and that expressions of different emotional states may have specific temporal patterns as well as morphological ones. For example, it has been shown [8] that when dynamic expressions were presented to participants, anger is recognized best when played at a medium speed, sadness is best recognised when it unfolds slowly, and happiness and surprise are recognised most accurately when the expressions change/develop quickly.

This mismatch between theory and reality is of crucial importance for researchers in the field of affective computing. Arguably, it might be possible to construct systems that can deliver believable emotional performances based on the posed prototypical patterns – after all, humans appear to have the capacity to be emotionally engaged by cartoon characters, line drawings and even inanimate objects. However, if the prototypical patterns do not reflect natural facial behaviour, then automatic systems trained on such patterns could not conceivably detect with any degree of accuracy the everyday emotions expressed by real people. The obvious answer would be to train perceptual systems on examples of naturally expressed emotion. However, it has been pointed

- I. Sneddon, M. McRorie, G. McKeown and J. Hanratty are with the School of Psychology Queen's University Belfast, Northern Ireland, UK.
E-mail: I.Sneddon@qub.ac.uk

out [9], [10] that there is a shortage of useful databases that can serve this purpose.

In addition, the currently existing databases fall short in many necessary aspects. The Belfast Naturalistic Database has low levels of quality in many of the recordings and suffers from copyright issues limiting its broad dissemination. Often databases have sporadic and sparse periods of emotion (AMI [11], Canal9 [12]) rather than targeting emotional reactions with specific tasks. Many databases do not combine audio and visual components (Cohn-Kanade [13], MMI-Facial Expression Database [14], [15]). Others are aimed at conversational interactions introducing speech as a complicating factor in the expression of emotion (AMI [11], SEMAINE [16]). The extent to which databases are annotated is also an issue, extensive annotation greatly adds to the utility of a database.

Ideally a database of emotional expression would contain high quality examples of a wide range of spontaneous natural emotion accompanied by a wealth of information about situational variables, individual characteristics of the actor, the emotion being experienced and its intensity. Normally, the more natural and spontaneous the emotional expression, the harder it is to capture, the less information is available about the actor and the harder it is to reproduce the eliciting circumstances. The more information you collect about the actor and the more you exert control over the eliciting circumstances, the less natural and spontaneous is the emotional expression you are likely to capture. This incompatibility between ecological validity and control means that, for the foreseeable future, useful databases are likely to be compromises between spontaneity and rigid control. We argue that exerting control over the context results in relatively standardised tasks that allow us to collect information about the participant's emotional state. Yet, allowing the participant to react to that context in their own way, results in tasks that are engaging enough to yield interesting and useful examples of natural emotional expression. In addition to high quality audio-visual recordings of facial expressions of people in a range of emotional situations, the database described in this paper offers information on self-report of emotion and intensity, continuous trace style ratings of valence and intensity, the sex of the participant, the sex of the experimenter, the active or passive nature of the induction task and it gives researchers the opportunity to compare expressions from people from more than one culture. As far as we know this is a unique set of attributes.

2 THE DATABASE

The Belfast Natural Induced Emotion Database is intended to provide examples of mild to moderately strong emotionally coloured naturalistic responses to a series of laboratory based tasks. Each example is a short (5s to 60s in length) video recording with stereo sound. Each task

is selected to provide participants with a fixed context designed to elicit an emotional state, strong enough to reveal individual differences between encoders, but not strong enough to cause ethical concerns (the database includes recordings of participants expressing emotion – referred to as ‘encoders’ and emotional rating information from other participants – referred to as ‘decoders’). While the tasks take place in the relatively artificial setting of a laboratory, we describe the emotion induced as natural, following the definition of naturalistic as “of a kind that might have to be dealt with in an application. The contrast is with idealized material, which is generated to match someone’s conception of what an emotion should be like” [17] (p 3518).

The selected tasks vary in their effectiveness at inducing the expected or “target” emotion (as measured by emotion self-report) varying from only 35% of encoders reporting Frustration to 92% reporting Amusement. However, this is to be expected. Although the physical context provided by the researchers is constant, the encoder is not, and their emotional reaction to the fixed context may vary for a range of reasons including their appraisal of the situation [18]. The database is divided into three broad sets that represent a categorisation based on different chronological periods in which the stimuli were collected and differing research goals for collecting the stimuli. Table 1 provides a brief summary of the database contents.

While this organisation of the database may seem at first sight to be rather arbitrary, it has been divided up in this manner because the nature of the tasks, details of the annotation, and the trace methods used, have evolved to fit the research requirements for which the sets of recordings were made. Thus, while it may look as if recordings of the same task appear in different sets, they differ in some aspect(s) of the methodology or the information accompanying them. The research motivations and differences between the database sets are described below.

2.1 The Tasks

The Set 1 tasks were initially developed as stimuli for research into the individual differences that might influence human abilities to encode and decode emotional signals. Previous emotion elicitation methods have generally been validated by measuring their ability to induce a reliable level of experienced emotion (e.g. [19]). Our primary focus, on the other hand, was to generate a sufficient range of emotional expression that would reveal the possible influence of factors such as gender, personality and emotional intelligence. Task development was driven by the need, therefore, to quickly elicit mild to moderately strong emotional behaviour. All of the tasks were conducted in the social presence of an experimenter (either male or female, to investigate same-sex and opposite-sex interactions). Four of the tasks were active in nature (the emotion was elicited as a result of

TABLE 1
Summary of Database Contents

	Activity/Sociality	Emotions Targeted	No Clips	Clip Length	Participants	Location
Set 1	active/social active/social active/social active/social passive/social	Frustration Disgust Surprise Fear Amusement	570	5 to 30 seconds	70 Male 44 Female	Northern Ireland
Set 2	active/social active/social active/social passive/non-social passive/non-social passive/non-social passive/non-social	Disgust Surprise Fear Amusement Anger Disgust Sadness	650	5 to 60 seconds	37 Male 45 Female	Northern Ireland
Set 3	active/social active/social passive/social	Disgust Fear Amusement	180	30 to 180 seconds	30 Male 30 Female	Northern Ireland Peru

the encoder actively engaging with the task) and one of the tasks was passive (viewing an amusing film). Although this final task was passive in nature, in the case of amusement, the use of film had previously been shown to quickly elicit emotionally expressive behaviour [19]. Note that within each set, the recordings of encoders watching a film excerpt have a fixed timeline – each encoder on a task is recorded watching the same excerpt. This provides a degree of synchrony to the film-viewing experience that is not necessarily present in the active tasks, although of course the encoders are still likely to vary in their emotional appraisal of the situation.

The Set 2 tasks were selected to allow comparison of these new tasks with more traditional film elicitors that had previously been validated for their ability to induce discrete emotions [19]. Three of the active/social tasks from set 1 (disgust, fear, surprise) are included along with 4 film-viewing tasks (anger, sadness, disgust and amusement). In Set 2, all of the film-viewing was conducted with the experimenter in the room, but out of sight and engaged in other tasks. Thus the four film-viewing tasks in Set 2 can all be categorised as passive/non-social. Note that in the ‘amusement’ film-viewing task; although the same film is viewed in both sets, the Set 2 recordings are of encoders watching a different and longer excerpt from that film.

The research motivation for Set 3 was to begin to explore cross-cultural differences in emotion encoding/decoding. The tasks for Set 3 were variants of the disgust and fear (both active/social) tasks and the amusement (passive/social) task from Set 1. While the process of sending emotional stimulus material to different countries for decoding is relatively straightforward, the process of collecting samples of encoded emotion from different countries is time consuming and resource intensive. The samples from Peru and Northern Ireland are therefore the first contributions to what we hope will become a more extensive resource as researchers from around the world add local examples.

2.2 Self-report of emotion

There is currently no external objective measure of our internal world of emotional experience. If we want to know about the emotions people are feeling, we have to ask them or indirectly infer through physiological measures. Physiological measures can be intrusive and interfere with the natural expression of an emotion. However, self-reports also generate problems. Self-reported emotional states need to be treated with caution because, even at best, they cannot tell us the whole story [20]. People have been found to vary greatly in their self-reports of emotional experiences [21], [22] and there is evidence that both sex [23] and culture [24] can influence self-reports of the nature of emotion and its intensity. However, despite these caveats, self-report remains an important window on the internal emotional world of the individual. Self reports of experienced emotion are available for each recording in the database but the methods used for measuring it vary across the sets. The first recordings made were of the female encoders in Set 1. Mindful of Russell’s [1] criticisms of the use of forced-choice selection from short lists of possible emotion terms, we adopted a completely open-ended free response method (no words were suggested and encoders could respond with as many descriptors as they wished). Intensity ratings for each descriptor were also collected. While this method yielded a rich and varied set of responses, and most were recognisable as single emotional states, it also produced descriptors that were not obviously emotional. For this reason, when male encoders were recorded, they were asked to select their emotional descriptions from a previously published list [25] but were again allowed to select as many as they wanted and were free to add any other descriptions not on the list. Intensity ratings for each descriptor were again collected.

Because the Set 2 recordings were collected to allow comparisons between methods for inducing emotion, it was important that our emotion self-reports were comparable with published validation measures for the film stimuli. We therefore adopted a variant of the procedure

described in [19] in which encoders are given a list of emotion words (in our case 10) and asked to rate the intensity of each one.

The method used to measure self-report emotion for the Set 3 recordings is similar to that described for Set 2 except that for the Peruvian encoders the 10 emotion words were translated into Spanish (the original English and Spanish versions are available for download with the database). In addition, because of the research focus on cross-cultural differences in encoding/decoding emotional signals, the Set 3 recordings are accompanied by self-report information about the encoder's experience of emotional signals from other cultures.

2.2.1 Rating of the emotion expressed in the recordings

Many of the recordings in Sets 1 and 3 and all of the recordings in Set 2 have been continuously rated by decoders using variants of a computer based tracing tool called 'FeelTrace' [26]. For Set 1 these ratings focus on the emotional dimensions of Intensity and Valence. Ratings of emotional intensity are a measure of the strength of the emotion the person in the recording (the encoder) is judged to be feeling irrespective of the precise nature of that emotion. Ratings of valence are a measure of the overall level of positive or negative emotion the encoder is judged to be feeling, again irrespective of the precise nature of the emotion [25]. Because the research focus behind making the Set 1 recordings was the investigation of individual differences in the encoding/decoding of emotional signals, the ratings for these clips are from relatively large numbers of inexperienced decoders ($N = 6$ to 258).

In Set 2 all of the recordings have been rated for Valence by a single experienced rater (with appropriate checks for inter-rater and intra-rater reliability detailed below in section 4.3).

In Set 3 the intensity of emotion is rated, but in this case, decoders are instructed to rate only the intensity of the target emotion. The research motivation for the Set 3 recordings was to investigate cultural differences in emotion signal encoding and decoding. The previous literature in this area has focussed strongly on a recognition paradigm based on discrete emotions. To allow direct comparison with previous studies (e.g. [27]) that have measured the perceived intensity of discrete emotions, we have modified the nature of the Trace annotations for this set of recordings so that, for instance in the 'disgust' task, raters were asked only to rate the intensity of disgust shown.

2.3 Ethical Approval

All participants who appear in the recordings have given written permission for the recorded images to be distributed for use by the wider research community. They were also asked about their willingness to have their images used to publicly disseminate research findings based on the recordings. Not all of the participants gave

permission for this and recordings in the database are labelled accordingly. Ethical approval for the induction methods and procedures involved in all recordings was granted by the Psychology Research Ethics Committee of Queen's University Belfast.

3 SET 1

This database consists of 570 video recordings (total length 237 minutes) of the face and torso of both male and female encoders performing a series of 5 emotion inducing tasks (Frustration, Disgust, Fear, Surprise and Amusement).

3.1 General Procedure

An opportunistic sample of 114 (44 female, 70 male) mainly undergraduate student encoders were recruited. Prior to recording and before being asked for their consent, encoders were given an information sheet that described the experimental set up and explained that their performance on the tasks would be recorded. Each encoder was recorded individually with only a single experimenter present. Approximately half of the male encoders and the female encoders were recorded by a male experimenter with the other half being recorded by a female experimenter. All clips in the database are accompanied by details of the sex of the experimenter. Although encoders were made explicitly aware of the camera before the session started, the experimenter implied that the primary interest was in their performance on the tasks. In the debriefing conversations at the end of recording, it was clear that most encoders had quickly forgotten the camera as they became absorbed in the tasks.

3.2 Recording Details

The digital video camera (a Panasonic NV-GS500) was placed approximately 2 metres directly in front of the encoder and 70cm from the ground allowing a view of the head and upper torso. Recording was on mini DV tape cassette and images were captured using Adobe Premiere in DV AVI Type 2 format with a screen resolution of 720 x 576 pixels.

3.3 Self Report Emotion

For Set 1 recordings, self report questionnaires were completed immediately following the final task. Female encoders were given an entirely free choice of the words they could use to describe their emotion and the number of emotions they could choose to describe. Encoders used between one and four words to describe their emotions. The overwhelming majority of words were recognisable as referring to a single emotional state (e.g. worry, disgust, calm), but a few carried less obvious emotional overtones (e.g. weird, silly, control). Each reported emotion is accompanied by an intensity rating

on a 0-9 scale anchored with the terms “Absolutely no emotion” and “Extreme emotion.” Because the female encoders had included non-emotional words in their self reports, it was decided to take a more prescriptive approach for the male self reports. Male encoders were given a list of emotion words (based on a previously published list [25]) to use when reporting the emotion(s) they had experienced. Again they were allowed to choose as many as they wished and could write their own emotion words if they felt the list did not cover the emotion(s) they had experienced. Almost all of the descriptions are recognisable as referring to a single main emotion but some encoders offered additional descriptions (e.g. ennui, giddy, tricked) while a very few offered a non-emotional commentary (e.g. surreal). Each reported emotion is accompanied by an intensity rating on a 0-9 scale anchored with the terms “Absolutely no emotion” and “Extreme emotion”. Because of the research focus on individual differences, each encoder performed the tasks in the same order. This order of presentation is almost certain to have influenced the responses of some of the encoders to some of the tasks. For instance, feelings of calm or relaxation reported for the amusement film task may, at least in part, have been due to the knowledge that the experimental session was coming to an end.

Details of the results for the self-reported emotion and intensities are available for each recording.

3.4 FeelTrace Annotation

Trace style continuous ratings have been made by decoders for a number of the recordings in this database. Ratings of emotional Intensity are available for 80 of the recordings; ratings of Valence are available for 58; and ratings of both are available for 37. The number of decoders that have provided ratings for each of the recordings varies from 6 to 258. Each rated recording is accompanied by the average rating and the number of decoders contributing to that rating is clearly indicated.

3.5 Induction Tasks

Encoders were seated at a small table in the presence of the experimenter and a series of tasks was presented to them one at a time and always in the order presented below.

3.5.1 Frustration Task

The Set 1 Frustration Task was a wire tracking task intended to induce emotion that encoders would describe in terms of the “anger family of words” [28] that includes frustration/irritation/anger/annoyance. This involved the encoder moving a copper ring (diameter = 25mm) attached to a thin rod held between the index finger and thumb along a zigzag bare copper wire without letting the ring touch the wire (a loud buzzer indicated an “error”). Encoders were instructed to move the ring from its insulated resting cradle at one end to the cradle

at the other end and back again as quickly as possible while making as few errors as possible. The pattern of bends in the wire was designed to make the task very difficult indeed with the intent of increasing frustration and/or annoyance. The first 30 seconds of their performance was recorded. The majority of the males (54%) reported having experienced the target emotion(s) with the others reporting either a positive experience (interest/humour/enjoyment) or a different negative experience (stress/anxiety/worry). Only 6% of the females reported experiencing the target emotion(s) with the overwhelming majority (80%) describing their emotional experience as worried/anxious/fearful/nervous. *Recording length*: 30 seconds – active/social task.

3.5.2 Disgust Task

The Set 1 Disgust Task introduced the encoder to the first of several black boxes (32cm x 23cm x 20cm) with a 10cm diameter hole cut in the top. This task was intended to induce disgust and inside this first box was a dish containing cold, cooked and cut spaghetti in sauce. The box was positioned so that the contents were not visible through the hole. Encoders were given a pen and sheets of blank paper were secured in front of them on a copy holder (see Figure 1). The encoder was instructed to reach inside the box with their non-dominant hand while using their dominant hand to write down as many words as possible to describe what was in the box. Recordings cover 30 seconds from the encoder putting their hand into the box.

The majority of males (59%) reported experiencing disgust. Almost all of those not reporting disgust reported positive emotion in the form of some combination of amusement/surprise/interest. Only a minority of women (46%) reported the target emotion with those not reporting disgust split fairly evenly between those reporting a positive experience (amusement/surprise/curiosity) and a negative experience (worried/nervous/apprehensive). *Recording length*: 30 seconds – active/social task.

3.5.3 Fear Task

The Set 1 Fear Task involved the experimenter carefully placing another black box in front of the encoder. This task was intended to induce fear in the encoder. The experimenter moved slowly and spoke quietly while moving and touching the box which had a hinged metal grill on top and was covered with several warning icons and images of spiders. Encoders were told that a sweet had been hidden in the box and their task was to reach in and find the sweet. The box was filled with finely shredded paper. The experimenter then began recording and the grill was carefully lifted back by the experimenter. If the encoder asked what was in the box they were told that “it will not harm you” or that there was “nothing in the box that will harm you”. Encoders were recorded for 30 seconds before being told that there was nothing in the box but the shredded paper. Among

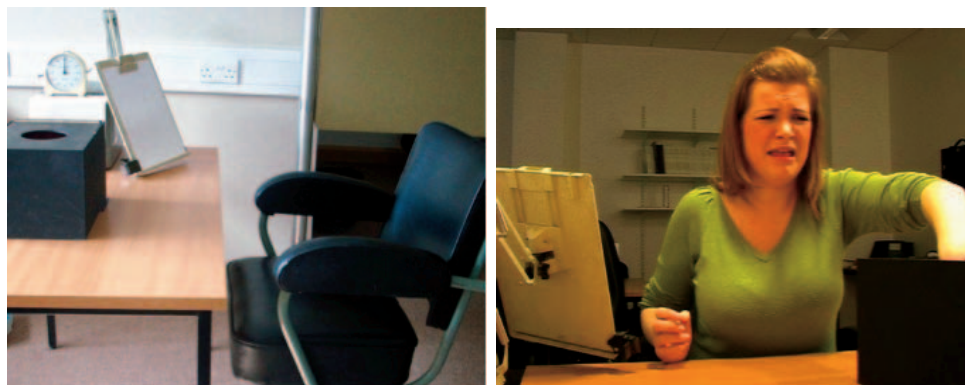


Fig. 1. The Set 1 & Set 2 Disgust Task and a still image from the Set 1 Disgust task

male encoders 50% reported feeling fear/anxiety/doubt with the majority of the other half of the sample reporting frustration and/or annoyance. A small number reported a positive experience (happy/calm). A higher proportion of the female encoders (66%) reported feeling fear/anxiety/doubt with the rest being split almost equally between a positive experience (amused) and a negative one (confusion or irritation). *Recording length*: 30 seconds – active/social task.

3.5.4 Surprise Task

The Set 1 Surprise Task was intended to induce surprise in the encoders and began with a third black box being placed in front of the encoder. As with the previous boxes there was a 10cm diameter hole in the top and encoders were asked to place their hand in the box to describe out loud what they found in the box. When encoders placed their hand in the box a photocell beam was broken and this in turn set off a loud noise (a personal attack alarm). Recordings cover five seconds from the point at which the encoder reached forward to place their hand in the box. Almost all of the male encoders (95%) reported feeling surprise/shock with the few who did not reporting some combination of fear/anxiety/stress/annoyance. Although the majority of women (62%) reported the target emotion, a substantial minority reported their emotion as some combination of fear/panic/anxiety. *Recording length*: 5 seconds – active/social task.

3.5.5 Amusement Task

The Set 1 Amusement Task was intended to induce amusement in encoders who were asked to watch a short video clip played on a monitor positioned 2 metres directly in front of them at a height of 70cm. The clip consisted of a 2 minute 31 second extract from an episode of Father Ted - a well known television comedy series in the UK and Ireland [29]. The 30 second recordings in the database show encoders watching the same segment of the film, the start-point being selected on the basis of a cue on the soundtrack. Almost all of the males (94%) reported feeling amusement/happiness with the

remaining few reporting either relaxation or boredom. The overwhelming majority of female participants (88%) also reported amusement/happiness with the remainder reporting some combination of relaxation/calm/relief. *Recording length*: 30 seconds – passive/social task.

4 SET 2

The defining characteristic of Set 2 was gathering instances of natural emotion with respect to the research goal of assessing the differences between active and passive forms of emotion induction. By active we mean emotion induction techniques in which the encoder is engaged and interacting as part of the task, passive refers to induction techniques where the encoder simply observes stimuli selected to induce emotion. This database consists of 650 video recordings (total length 458 minutes) of the face and torso of both male ($n=37$) and female ($n=45$) encoders performing a series of three active emotion inducing tasks (Disgust, Surprise and Fear) and four passive emotion induction tasks (Disgust, Sadness, Anger and Amusement). This second set of recordings was based on a different mixture of tasks from Set 1. The tasks employed in this set used active induction techniques to elicit disgust, fear and surprise and passive induction (viewing film) to elicit disgust, sadness, anger and amusement.

4.1 General Procedure

An opportunistic sample of ninety encoders (48 female, 42 male) was recruited. The majority (69, 44 female) were undergraduate students. The remaining 21 were either postgraduate students (6 male, 3 female) or employed professionals (11 male, 1 female). The mean age of participants was 23.78 years ($SD=6.88$; Female; $M=23.35$ $SD=7.31$, Male; $M=24.26$, $SD=6.41$). Of these participants 82 (37 male) consented to inclusion of their video recordings in this database. The remaining general procedures and recording details were as described for Set 1.

4.2 Self Report Emotion

Self-reports of the emotion experienced during each of the active tasks were collected on completion of the third active task and before starting the passive tasks. Encoders rated their experience of each ten emotion terms on a 0-9 scale anchored with the terms “none of that emotion” and “extreme emotion.” The emotion terms rated were; afraid, amused, angry, anxious, disgusted, relaxed, happy, interested, sad and surprised. Details of the results for the self-reported emotion and intensities are available for each recording.

4.3 FeelTrace Annotation

Trace style continuous ratings of Valence have been made for all of the recordings in this set by one experienced female rater and a global mean averaged over time is provided for each recording. To assess rater reliability 10% of the clips were rated twice by the main rater and a measure of intra-rater reliability was calculated (by correlating the two traces for each clip). The average correlation between the first and second ratings was $r = .82$ (average correlation for each clip type ranged from .68 to .92) indicating that the main rater was consistent in her ratings. In addition, two experienced raters (one male one female) rated another 10% of the clips and inter-rater reliability was calculated by correlating the traces for rater 1 (the main rater) with those of rater two and three. The average correlation between raters’ traces was .58. Intra class correlations (ICCs) were also conducted on the range of each trace and the sum of positive and sum of negative ratings. ICCs using a two-way mixed model to assess absolute agreement between raters indicated that rater agreement was high. ICC for range of emotional behaviour, sum of negative ratings and sum of positive ratings was .82, .91 and .93 respectively. Although the raw correlations between the three raters traces was relatively low, the ICCs indicated that there was substantial agreement between raters on range, sum of negative and sum of positive emotional behaviour yielding an acceptable inter rater reliability.

4.4 Induction Tasks

Encoders were seated at a small table and a series of tasks was presented to them one at a time and always in the order presented below. The active induction tasks were the three tasks designed to elicit disgust, surprise and fear that are described in “Set 1.” After completing these three tasks encoders were given the emotion self-report questionnaire asking them to rate the emotion they had experienced during each task. The passive induction tasks involved the encoders being recorded while watching a series of film clips. A laptop computer, placed on the table in front of the encoder, was used to show the film extract used in the Set 1 Amusement task and three additional films, previously reported to induce feelings of sadness, disgust and anger [19].

4.4.1 Active Induction Tasks

As in Set 1, the tasks were designed so that the focus of the encoder would be on completing the tasks in accordance with the instructions given. Encoders were presented with the disgust task first followed by the surprise task and finally the fear task. Details of the procedure are as described in Set 1. Self reports of the mean target emotion experienced are provided in Table 2. Database recordings are 30 seconds in length for the disgust and fear tasks and 5 seconds in length for the surprise task. All three tasks can be categorised as active/social.

4.4.2 Passive Induction Tasks

During the film presentation stage of testing, encoders were seated at the same table as before with the camera in the same position. A laptop computer was positioned in front of the encoder on the table. Encoders were informed that they would see some scenes of violence, a surgical procedure and bereavement and were reminded that should they feel uncomfortable or did not want to continue watching any of the clips they could press any key on the keyboard and skip to the introduction slide for the next induction film. Each film clip was preceded by a neutral film clip (tranquil nature scenes) to induce a neutral affective state and to avoid any carry over effects between clip presentations. These clips were taken from “The Blue Planet” [30] and depict various scenes of fish swimming in the ocean. Each clip was 30 seconds in length and the clips were validated using a panel of 30 student decoders and shown to induce a neutral affective state.

The sadness inducing clip came from “The Champ” [31] and depicted a young boy crying at the death of his father. This clip has been shown to elicit sadness [19]. The disgust clip is a non-commercial amputation film which depicts surgery to amputate a human forearm and can be freely downloaded from <http://spl.stanford.edu/movs/surgery.mov>. This clip has been shown to reliably induce disgust [19]. The anger inducing clip came from the film “Cry Freedom” [32] and depicted a crowd of peaceful protesters being indiscriminately shot and killed by police. This clip has been shown to elicit anger [19]. Finally, the amusing clip was the same as that used in the Set 1 amusement task but note that the recorded sequence is the final 60 seconds of the clip and not the 30 second extract used for the Set 1 recording.

Participants watched the clips alone. The experimenter remained in the room out of sight behind a partition and told participants that she would be doing other unrelated work while they watched the clips. To begin, participants were presented with the following onscreen instructions “This part of the experiment is testing your ability to ‘get into’ a film clip. Please watch each clip carefully. Click the left mouse button to continue”. The clips were presented in the same order each time; neutral clip, sad clip, neutral, disgust, neutral, anger, neutral, amusement.

TABLE 2
 Mean self reported levels of the target emotion and additional elicited emotions

Task	Target Emotion	Male	Female	Total	Other emotion elicited
Spaghetti	Disgust	3.55	4.23	4.19	Interest 5.62 Surprise 5.57 Amusement 5.02
Alarm	Surprise	6.74	6.91	7.60	
Imaginary Spider	Fear	3.45	4.92	4.60	Anxiety 5.15
Neutral	Relaxed	5.62	6.15	6.48	
Disgust Video	Disgust	4.40	6.00	5.76	
Sad Video	Sadness	4.76	6.52	6.23	
Anger Video	Anger	5.02	6.48	6.41	Sadness 5.49 Disgust 6.53
Amusing Video	Amusement	5.79	6.56	6.87	Relaxation 5.82 Happiness 6.58

The amusing clip was intentionally presented last so that encoders would be in a positive mood before leaving the lab. Following each of the target clips, the onscreen instructions asked encoders to rate their emotional experiences during the clip using the same scale as before. Means of encoder's self reported emotion in response to each of the database clips can be found in Table 2. Once encoders had watched all of the clips and completed the rating scales they were debriefed and informed of the purposes of the study. Database recordings are the final sixty seconds of the encoder viewing each film. All of the Set 2 film-viewing tasks can be categorised as passive/non-social.

5 SET 3

This section of the database was collected to allow comparison of emotional expression by people from two countries (Peru and Northern Ireland) representing quite different cultures. The need for such diversity in databases is often overlooked, it can result in it only being possible to address cross-cultural questions of emotion perception in one direction [33] and the dangers of relying too heavily on one culture have recently been amplified [34]. When compiling a database of human behaviour to be used as examples, it would be desirable for the examples to represent as broad a range of humanity as possible. A broad collection is important for many reasons, not least that potential emotion-oriented computer systems will work only poorly if provided with a narrow range of emotional examples [35]. Although the ages of participants fall within a relatively narrow range (and include only adults), the inclusion of males and females and at least some cultural diversity allows us to claim that we have begun the task of broadening our range of examples. This database consists of 180 video recordings (total length 90 minutes) of the face and torso of both Peruvian male (n=15) and female (n=15) and Northern Irish male (n=15) and female (n=15) encoders performing a series of three emotion inducing tasks (Disgust, Fear and Amusement). This third set of recordings was based on modified versions of three of the Set 1 Tasks (Disgust, Fear and Amusement). The encoders recorded for this set were recruited in Chincha Alta, Peru and in Belfast,

Northern Ireland to allow cross-cultural comparison of expression of emotion.

5.1 General Procedure

An opportunistic sample of 60 participants were recruited; 30 in Peru (15 female, 15 male) and 30 in Northern Ireland (15 female, 15 male). In the Peruvian sample the occupations were students, domestic, agrarian, and factory workers and the mean age of participants was 32.54 years; Female; M=32.38, Male; M=32.69). Participants in the Northern Ireland sample were all undergraduate students with a mean age of 20 years; Female; M=20, Male; M=20). The remaining general procedures for the Northern Ireland sample were as described for Set 1. The Peruvian encoders were recorded in an enclosed outdoor area. Encoders entered the study area which contained a table and seat where they were informed of the general nature of the tasks and consented to participate. Each encoder was recorded individually with in the presence of two experimenters, the first spoke Spanish as a second language and the second was a native Peruvian Spanish speaker and local to the area who ensured instructions were correctly understood. Due to the fact that recordings take place in an enclosed outdoor area there are times when external sounds are intrusive.

5.2 Recording Details

The camera (a Sony HDR-CX105E) was placed approximately 2 metres directly in front of the participant and 70cm from the ground allowing a view of the head and upper torso. Recording was on High Definition Full HD AVCHD and subsequently compressed using the H.264/MPEG-4 (x264) codec. Video was compressed to 7,949kbit/s (resolution 1920 x 1080 pixels).

5.3 Self Report Emotion

Self-reports of the emotion experienced during each of the active tasks were collected on completion of the third task using a similar questionnaire to that used for Set 2 in which the target words were translated into an equivalent or near equivalent word in Spanish



Fig. 2. Set 3 Participant PF01 Fear and Participant PF02 Amused

(available for download with the database). Details of the results for the self-reported emotion and intensities are available for each recording. Additionally a further questionnaire attempted to address the issue of exposure to other cultures and people from other cultures, asking people to assess the amount of time they have spent interacting with people from various parts of the world and their exposure to television and films from the same parts of the world. The details will be available for each recording.

5.4 Feeltrace Annotation

Trace style continuous ratings are available for 120 of the recordings in this database (40 encoders for each task: 10 Peru males; 10 Peru females; 10 N.Ireland males; 10 N.Ireland females). Each recording has been rated by 20 decoders using a variant of the rating methodology described for Sets 1 and 2. In this Set decoders are asked to continuously rate the intensity of the target emotion (Disgust, Fear or Amusement).

5.5 Induction Tasks

Encoders were seated at a small table and a series of tasks was presented to them one at a time and always in the order presented below. The induction tasks were modified versions of the three tasks designed to elicit disgust, fear and amusement that are described in “Set 1”.

5.5.1 Disgust Task

Procedural details for this task are as described for Sets 1 and 2 except that, instead of being asked to write down the words describing what was in the box, encoders were instructed to verbally describe the box’s contents using as many descriptive words as possible. The writing procedure in previous versions often led to the encoders looking away from the camera, the verbal procedure was an attempt to minimise this problem. Additionally, the Peruvian encoders are standing behind the table in this version rather than seated. *Recording length*: 30 seconds – active/social task.

5.5.2 Fear Task

Procedural details are as described for Sets 1 and 2. With the exception that in the Peruvian version the encoders are standing behind the table rather than seated and the dimensions of the boxes used were slightly larger. Warning signs were attached to the sides in Spanish stating “Cuidado” and “Peligro” (“Watch out”, and “Danger”) and there were two visible images of spiders attached to the box. *Recording length*: 30 seconds – active/social task.

5.5.3 Amusement Task

Procedural details for the Northern Ireland sample are as described for Set 1. The Northern Irish encoders are recorded while viewing the same 30 second “Father Ted” excerpt used in Set 1 [29] with the start point selected on the basis of the same cue on the soundtrack. The Peruvian encoders, however the use of the English language film clip with its culture-specific humour content was not appropriate.

We selected an alternative film clip that pilot testing indicated was likely to be found humorous by the majority of Peruvian encoders. The clip was “Mr Bean – At the Dentist,” 2 minutes 35 seconds [36]. The 30 second recordings in the database show each of the Peruvian encoders watching the same segment of the film, the start-point being selected on the basis of a cue on the soundtrack. *Recording length*: 30 seconds – passive/social task.

6 AVAILABILITY

The Belfast Induced Natural Emotion Database is made freely available to the research community. It is available through a web-accessible interface with url <http://www.psych.qub.ac.uk/BINED>. Those wishing to access the database are required to sign an End User Licence Agreement.

7 CONCLUSION

Development of effective emotion-oriented computer systems is dependent on access to examples of natural

emotional behaviour. The recordings comprising this database were collected initially to facilitate psychological investigation of sex differences, cultural differences, and individual differences in the expression and perception of emotion. However, we also appreciate the value of this database to researchers in the field of affective computing. This unique collection of mild to moderate natural emotional behaviour with a range of useful annotation should prove an invaluable resource for those working on the perception and recognition of emotion. The eliciting situations offer a compromise between entirely natural situations where the encoder is unaware of being observed (with the inevitable practical and ethical problems this entails) and situations in which the experimenter exerts complete control over encoder facial expressions by prescribing muscle configurations. In the eliciting procedures described here, the situations are fixed but encoders are free to respond as they see fit. Of course, in any laboratory based experimental setting, the extent to which an individual's reactions are entirely 'spontaneous' is open to question. We have tried to create sufficiently engaging tasks that would retain an encoder's attention. Inevitably, the extent to which the encoder was led to believe the focus of the research was on something other than the facial expression of emotion differed from task to task. There may be ways to investigate the degree of conscious control that encoders exhibit in their behavioural responses to the different tasks but there is probably no way to ensure that behaviour in a laboratory setting exactly mirrors that in a more natural setting. However, we should remember that even in entirely natural social interactions, the extent to which an individual consciously manages the impression they give varies from person to person; from situation to situation and probably from moment to moment.

These eliciting situations, where the tasks are presented to encoders in the same order, were designed to illuminate individual variations in encoder responses to the situations. Each individual's emotional appraisal of a situation varies for a wide variety of reasons and our recent research has begun to ask what proportion of that variance can be attributed to predictors such as sex, personality or cultural background. Such variation is, of course, an important facet of human behaviour and we have barely begun to explore how much such variation will need to be addressed in artificial systems designed to encode or decode human emotional signals. Questions about the relative importance of universals in the expression and perception of human emotion are empirical questions and it is vital that affective computing does not make premature assumptions about the answers. It is important that ongoing research in this field has access to examples of relatively natural emotional behaviour performed by a range of people of differing genders, cultures and personalities. The database we describe in this paper at least makes a start by providing some of the required variation in emotional responses. However, the database is by no means a finished product and should

be regarded rather as a 'work in progress'.

We are continuing to collect recordings of emotional behaviour as part of our ongoing psychological research on different aspects of emotion expression and perception and we intend to expand this database as these recordings and their annotations become available.

ACKNOWLEDGMENTS

The authors would like to acknowledge the assistance of Roberto Sebastian Castilla for translation and support of the recording in Peru, and to thank all the participants who gave their up their time and agreed to the inclusion of their videos in the database.

REFERENCES

- [1] J. A. Russell, "Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies," *Psychol Bull.*, vol. 115, no. 1, pp. 102–41, 1994.
- [2] P. J. Naab and J. A. Russell, "Judgments of emotion from spontaneous facial expressions of New Guineans," *Emotion*, vol. 7, no. 4, pp. 736–744, 2007.
- [3] P. Ekman and W. V. Friesen, "Facial Action Coding System," p. 233, 1978.
- [4] J. Carroll and J. Russell, "Facial expressions in Hollywood's portrayal of emotion," *J Pers Soc Psychol*, vol. 72, pp. 164–176, 1997.
- [5] K. Scherer and H. Ellgring, "Are facial expressions of emotion produced by categorical affect programs or dynamically driven by appraisal," *Emotion*, vol. 7, no. 1, pp. 113–130, 2007.
- [6] A. Batliner, K. Fischer, R. Huber, J. Spilker, and E. Nöth, "How to find trouble in communication," *Speech communication*, vol. 40, no. 1-2, pp. 117–143, 2003.
- [7] K. Edwards, "The face of time: Temporal cues in facial expressions of emotion," *Psychological Science*, vol. 9, no. 4, p. 270, 1998.
- [8] M. Kamachi, V. Bruce, S. Mukaida, J. Gyoba, S. Yoshikawa, and S. Akamatsu, "Dynamic properties influence the perception of facial expressions," *Perception*, vol. 30, pp. 875–887, 2001.
- [9] R. Cowie, E. Douglas-Cowie, and C. Cox, "Beyond emotion archetypes: databases for emotion modelling using neural networks," *Neural Netw*, vol. 18, no. 4, pp. 371–88, 2005.
- [10] R. Cowie, E. Douglas-Cowie, J.-C. Martin, and L. Devillers, "The essential role of human databases for learning in and validation of affectively competent agents," in *A Blueprint for Affective Computing: a Sourcebook and Manual*, K. R. Scherer, T. Bänziger, and E. Roesch, Eds. Oxford: Oxford University Press, 2010, pp. 151–165, in print.
- [11] I. Mccowan, G. Lathoud, M. Lincoln, A. Lisowska, W. Post, D. Reidsma, and P. Wellner, "The AMI meeting corpus," in *Measuring Behavior 2005, 5th International Conference on Methods and Techniques in Behavioral Research*, L. L. L.P.J. Noldus, F. Grieco and P. Zimmerman, Eds., Wageningen, 2005.
- [12] A. Vinciarelli, A. Dielmann, S. Favre, and H. Salamin, "Canal9: A database of political debates for analysis of social interactions," 2009.
- [13] T. Kanade, J. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," in *Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, 2000, pp. 46–53.
- [14] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, "Web-based database for facial expression analysis," in *Proceedings of IEEE Int'l Conf. Multimedia and Expo (ICME'05)*, Amsterdam, The Netherlands, July 2005, pp. 317–321.
- [15] M. Valstar and M. Pantic, "Induced disgust, happiness and surprise: an addition to the mmi facial expression database," in *Proceedings of Int'l Conf. Language Resources and Evaluation, Workshop on EMOTION*, Malta, May 2010, pp. 65–70.
- [16] G. Mckeown, M. Valstar, R. Cowie, and M. Pantic, "The semaine corpus of emotionally coloured character interactions," in *Proceedings of IEEE Int'l Conf. Multimedia, Expo (ICME'10)*, Singapore, July 2010, pp. 1079–1084.

- [17] R. Cowie, "Perceiving emotion: towards a realistic understanding of the task," *Philos T R Soc B*, vol. 364, no. 1535, pp. 3515–3525, 2009.
- [18] K. R. Scherer, "Appraisal theory," in *Handbook of Cognition and Emotion*, T. Dalgleish and M. Power, Eds. Chichester, UK: Wiley, 1999, pp. 637–663.
- [19] J. Rottenberg, R. Ray, and J. Gross, "Emotion elicitation using films," in *Handbook of emotion elicitation and assessment*, J. A. Coan and J. J. B. Allen, Eds. Oxford: Oxford University Press, 2007, pp. 9–28.
- [20] J. Russell and L. Barrett, "Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant," *J Pers Soc Psychol*, vol. 76, no. 5, pp. 805–819, 1999.
- [21] L. Barrett, "Feelings or words? Understanding the content in self-report ratings of experienced emotion," *J Pers Soc Psychol*, vol. 87, no. 2, p. 266, 2004.
- [22] M. D. Robinson and G. L. Clore, "Belief and feeling: Evidence for an accessibility model of emotional self-report," *Psychological Bulletin*, vol. 128, no. 6, pp. 934–960, 2002.
- [23] L. F. Barrett, L. Robin, P. R. Pietromonaco, and K. M. Eysell, "Are women the 'more emotional' sex? Evidence from emotional experiences in social context," *Cognition and Emotion*, vol. 12, no. 4, pp. 555–578, 1998.
- [24] H. A. Effenbein and N. Ambady, "Cultural similarity's consequences: A distance perspective on cross-cultural differences in emotion recognition," *Journal of Cross-Cultural Psychology*, vol. 34, no. 1, pp. 92–110, 2003.
- [25] L. Devillers, R. Cowie, J. Martin, E. Douglas-Cowie, S. Abrilian, and M. McRorie, "Real life emotions in French and English TV video clips: an integrated annotation protocol combining continuous and discrete approaches," *5th international conference on Language Resources and Evaluation (LREC 2006)*, Genoa, Italy, 2006.
- [26] R. Cowie, E. Douglas-Cowie, S. Savvidou, E. McMahon, M. Sawey, and M. Schroder, "'Feeltrace': An instrument for recording perceived emotion in real time," in *Proceedings of the ISCA Tutorial and Research Workshop on Speech and Emotion*, 2000.
- [27] P. Ekman, W. V. Friesen, M. O'Sullivan, A. Chan, I. Diacoyanni-Tarlatzis, K. Heider, R. Krause, W. A. LeCompte, T. Pitcairn, and P. E. Ricci-Bitti, "Universals and cultural differences in the judgments of facial expressions of emotion," *J Pers Soc Psychol*, vol. 53, no. 4, pp. 712–7, 1987.
- [28] P. Ekman, "Strong evidence for universals in facial expressions: A reply to Russell's mistaken critique," *Psychol Bull*, vol. 115, pp. 268–268, 1994.
- [29] H. Baker, P. Shortt, G. P. Perkins, and D. D. Lowney, "Father Ted, episode 1 series 2, 'Hell' [situation comedy]," Hat Trick Productions/Channel 4, United Kingdom, 1996.
- [30] A. Fothergill, A. P. Byatt, and A. D. Fothergill, "The Blue Planet, [television documentary]," BBC, 2001.
- [31] D. P. Lovell and F. D. Zeffirelli, "The Champ, [motion picture]," MGM/Pathé Home Video, United States, 1979.
- [32] N. Spencer, J. P. Briley, and R. D. Attenbrough, "Cry Freedom [motion picture]," MCA/Universal Home Video, United States, 1987.
- [33] I. Sneddon, G. McKeown, M. McRorie, and T. Vukicevic, "Cross-cultural patterns in dynamic ratings of positive and negative natural emotional behaviour," *PLoS ONE*, vol. 6, no. 2, p. e14679, 2011.
- [34] J. Henrich, S. J. Heine, and A. Norenzayan, "The weirdest people in the world?" *Behavioral and Brain Sciences*, vol. 33, no. 2-3, pp. 61–83, 2010.
- [35] I. Sneddon, P. Goldie, and P. Petta, "Ethics in emotion-oriented-systems: the challenges for an ethics committee," in *Emotion-Oriented Systems: The Humaine Handbook*, P. Petta, C. Pelachaud, and R. Cowie, Eds. Springer, 2010.
- [36] J. Birkin, P. Weiland, P. Bennett-Jones, and S. Vertue, "The Trouble with Mr Bean," 1991. [Online]. Available: <http://www.youtube.com/watch?v=IW6R9kSGV2Q>



perception/expression. He was on the board of management of the HUMAINE network of excellence and is a member of the SEMAINE project.



in the encoding and decoding of emotion. Her other research interests include emotion induction techniques, and cross-cultural differences in expressing and perceiving emotion. She is a member of the SEMAINE project.



cultural emotion perception. He is a member of the SEMAINE project.



Ian Sneddon graduated with BSc in Psychology in 1974 followed by a PhD in Animal Social Behaviour, both from St Andrews University. He is currently a senior lecturer in the School of Psychology, Queen's University Belfast. Moving from ethological study of animal behaviour to the detailed observation of human behaviour, he is a certified FACS coder. Research interests include the expression and perception of emotion; emotion induction techniques and ethical issues; cross-cultural differences in emotion

Margaret McRorie Margaret McRorie studied Psychology at Queen's University Belfast, graduating with a BA in 1997, followed by a PhD in 2001. She was appointed as Individual Differences lecturer in the Psychology Department at Queen's in 2002. Her research within the area of Individual Differences has continued, and she is currently focusing on individual differences issues within the area of emotion research. She is a certified FACS coder, and is specifically interested in exploring individual characteristics

Gary McKeown Gary McKeown is a cognitive psychologist at the School of Psychology, Queen's University Belfast. His PhD explored mechanisms of implicit learning in the control of complex systems. His research focuses on communication, with interest in risk perception and decision making in environmental and health settings. This led to an interest in emotion and in particular the inter-relationship of cognition and emotion. Recent research has focused on emotion, social signal processing and the cross-

Jennifer Hanratty Jennifer Hanratty received her BSc in Psychology in 2003 from Queen's University Belfast (QUB). She returned to QUB in 2006 and was awarded her PhD on individual and situational differences in emotional expression in 2010. She is currently working as research fellow on the SEMAINE project. Her research interests include natural emotional expression and perception, emotion regulation and personality processes related to emotion.